Nicolae Sfetcu

Etica Big Data în cercetare

Colecția ESEURI

MultiMedia Publishing

Etica Big Data în cercetare

Nicolae Sfetcu

06.07.2019

Sfetcu, Nicolae, "Etica Big Data în cercetare", SetThings (6 iulie 2019), DOI: 10.13140/RG.2.2.27629.33761, MultiMedia Publishing (ed.), ISBN: 978-606-033-228-2, URL = https://www.telework.ro/ro/e-books/etica-big-data-in-cercetare/

Email: nicolae@sfetcu.com

Abstract

Principalele probleme cu care se confruntă oamenii de știință în lucrul cu seturile mari de date (Big Data), evidențiind principale aspecte etice, luând în considerare inclusiv legislația din Uniunea Europeană. După o scurtă *Introducere* despre Big Data, secțiunea *Tehnologia* prezintă aplicațiile specifice în cercetare. Urmează o abordare a principalelor probleme filosofice specifice în *Aspecte filosofice*, și *Aspecte legale* cu evidențierea problemelor etice specifice din Regulamentul UE privind protecția datelor 2016/679 (General Data Protection Regulation, "GDPR"). Secțiunea *Probleme etice* detaliază aspectele specifice generate de Big Data. După o scurtă secțiune *Cercetarea Big Data*, finalizez lucrarea cu prezentarea *Concluziilor* pentru etica cercetării în lucrul cu Big Data.

1. Introducere

Termenul Big Data se referă la extragerea, manipularea şi analiza unor seturi de date care sunt prea mari pentru a fi tratate în mod obişnuit. Din această cauză se utilizează software special şi, în multe cazuri, şi calculatoare şi echipamente hardware special dedicate. În general la aceste date analiza se face statistic. Pe baza analizei datelor respective se fac de obicei predicţii ale unor grupuri de persoane sau alte entităţi, pe baza comportamentului acestora în diverse situaţii şi folosind tehnici analitice avansate. Se pot identifica astfel tendinţe, necesităţi şi evoluţii comportamentale ale acestor entităţi. Oamenii de ştiinţă folosesc aceste date pentru cercetări în meteorologie, genomică, (Nature 2008) conectomică, simulări fizice complexe, biologie, protecţia mediului , etc. (Reichman, Jones, and Schildhauer 2011)

Odată cu creşterea volumului de date pe Internet, în media socială, cloud computing, dispozitive mobile şi date guvernamentale, Big Data devine în acelaşi timp o ameninţare şi o oportunitate pentru cercetători în ceea ce priveşte gestionarea şi utilizarea acestor date, menţinând în acelaşi timp drepturile persoanelor implicate.

1.1 Definiţii

Big Data includ, de obicei, seturi de date cu dimensiuni care depăşesc capacitatea software şi hardware obişnuite, folosind date nestructurate, semi-structurate şi structurate, cu accentul pe datele nestructurate. (Dedić and Stanier 2017) Dimensiunile Big Data au crescut în timp din 2012, de la câteva zeci de terabyte până la multe exabyte de date. (Everts 2016) Eficientizarea lucrului cu Big Data implică învăţarea maşinilor pentru a detecta modele, (Mayer-Schönberger and Cukier 2014) dar adesea aceste date sunt un produs secundar al altor activităţi digitale.

O definiție din 2018 afirmă că "**Big Data** sunt datele care necesită instrumentele de calcul paralel pentru a gestiona datele", aceasta reprezentând o turnură în informatică, prin utilizarea teoriilor de programare paralelă și lipsa unor garanții presupuse de modelele anterioare." Big Data utilizează statistici inductive și concepte de identificare a sistemelor neliniare pentru a deduce legi (regresii, relații neliniare și efecte cauzale) din seturi mari de date cu densitate scăzută de informații pentru a obține relații și dependențe sau pentru a efectua predicții ale rezultatelor și comportamentelor.

La nivelul Uniunii Europene nu există o definiție obligatorie dar, în conformitate cu Avizul 3/2013 al Grupului european de lucru privind protecția datelor,

> "**Big Data** este un termen care se referă la creșterea enormă a accesului și a utilizării automate a informațiilor: se referă la cantitățile uriașe de date digitale controlate de companii, autorități și alte organizații mari, care sunt supuse unor analize ample bazate pe utilizarea de algoritmi. Big Data pot fi folosite pentru a identifica tendințele și corelațiile generale, dar pot fi utilizate și pentru a afecta direct persoanele." (European Economic and Social Committee 2017)

Problema cu această definiție e că nu ia în considerare reutilizarea datelor cu caracter personal.

Regulamentul nr. 2016/679 definește **datele personale** (articolul 4, paragraful 1) drept

> "orice informație referitoare la o persoană fizică identificată sau identificabilă (persoana vizată); o persoană fizică identificabilă este cea care poate fi identificată, în mod direct sau indirect, în special prin referire la un identificator cum ar fi un nume, un număr de identificare, date de localizare, un identificator online sau unul sau mai mulți factori specifici identității fizice, fiziologice, genetice, mentale, economice, culturale sau sociale a acelei persoane fizice."

Definiția se aplică, la nivelul UE, și persoanelor neidentificate dar care pot fi identificate prin corelarea datelor anonime cu alte informații suplimentare. Datele cu caracter personal, o dată anonimizate (sau pseudo-anonimizate), pot fi prelucrate fără a fi nevoie de o autorizație, ținându-se totuși cont de riscul re-identificării persoanei vizate.

1.2 Dimensiunile Big Data

Datele sunt partajate și stocate pe servere, prin interacțiunea dintre entitatea implicată și sistemul de stocare. În acest context, Big Data se poate clasifica în sisteme active (interacțiune sincronă, datele entității sunt trimise direct către sistemul de stocare), și sisteme pasive (interacțiune asincronă, datele sunt colectate printr-un intermediar și apoi introduse în sistem.

De asemenea, datele pot fi transmise direct în mod conștient, sau ne-conștient (dacă persoana ale cărei date sunt transmise nu este notificată la timp și clar). Datele sunt apoi prelucrate pentru a genera statistici.

În funcție de ținta analizelor statisticilor respective, dimensiunile datelor pot fi a) individuale (este analizat o singur entitate); sociale (se analizează grupuri discrete de entități din cadrul unei populații; și hibride (când o entitate este analizată prin prisma apartenenței sale la un grup deja definit).

Producția actuală imensă de date generate de utilizatori este estimată că va crește cu 2000% 1 2 3 la nivel mondial până în 2020, și sunt adesea nestructurate. (a7) În general, Big Data se caracterizează prin:

- Volum (cantitatea de date);
- Varietate (produse de diferite surse în diferite formate);
- Viteză (viteza de analiza online a datelor);
- Veracitate (datele sunt incerte și trebuie verificate);
- Valoare (evaluată prin analiză).

Volumul de date produse și stocate evoluează în prezent exponențial, peste 90% din ele fiind generate în ultimii patru ani. (European Economic and Social Committee 2017) Volumele mari necesită viteză mare de analiză, cu impact puternic asupra

veracității. Datele incorecte au potențialul de a genera probleme atunci când sunt folosite în procesul de decizie.

Una din probleme important cu Big Data este dacă este nevoie de datele complete pentru a trage anumite concluzii cu privire la proprietățile lor, sau este suficient un eșantion. Big Data conține chiar în nume un termen legat de dimensiune, care este o caracteristică importantă a Big Data. Dar eșantionarea (statistică) permite selectarea unor puncte corecte de colectare de date dintr-un set mai larg pentru a estima caracteristicile întregii populații. Big Data pot fi eșantionate pe diferite categorii de date în procesul de selecție a probelor cu ajutorul unor algoritmii de eșantionare pentru Big Data.

2. Tehnologia

Datele trebuie procesate cu instrumente avansate de colectare și analiză, pe baza unor algoritmi prestabiliți, pentru a putea obține informații relevante. Algoritmii trebuie să ia în considerare și aspecte invizibile pentru percepțiile directe.

În 2004 Google a publicat o lucrare despre un proces numit MapReduce care oferă un model de procesare paralelă. (Dean and Ghemawat 2004) De asemenea, MIKE2.0 este o aplicație în sursă deschisă pentru managementului informațiilor. (MIKE2.0 2019) Mai multe studii din 2012 au arătat că arhitectura optimă pentru a aborda problemele din Big Data sunt cele cu mai multe straturi. O arhitectură paralelă distribuită distribuie date pe mai multe servere (medii de execuție paralelă) putându-se îmbunătăţi astfel dramatic vitezele de procesare a datelor.

Conform unui raport al Institutului Global McKinsey din 2011, principalele componente și ecosisteme ale Big Data sunt: (Manyika et al. 2011) tehnici de analiză a

datelor (învățarea mașinilor, prelucrarea limbajului natural, etc.), tehnologii de mari dimensiuni (business intelligence, cloud computing, baze de date), și vizualizări (diagrame, grafice, alte afișări ale datelor).

Big Data furnizează informații în timp real sau aproape real, evitându-se astfel latența ori de câte ori este posibil.

2.1 Aplicații

Big Data în procesele guvernamentale cresc eficiența costurilor, productivitatea și inovația. Registrele civile sunt o sursă pentru Big Data. Datele prelucrate ajută în domenii critice de dezvoltare, cum ar fi îngrijirea sănătății, ocuparea forței de muncă, productivitatea economică, criminalitatea, securitatea și gestionarea dezastrelor naturale și a resurselor. (Kvochko 2012)

De asemenea, Big Data oferă o infrastructură care este permite evidențierea incertitudinilor, a performanței, și disponibilitatea componentelor. Tendințele și predicțiile în industrie necesită o cantitate mare de date și instrumente avansate de predicție.

Big Data contribuie la îmbunătățirea asistenței medicale prin furnizarea de medicamente personalizate și analize prescriptive, intervenții clinice cu evaluarea riscurilor și analize predictive, etc. Nivelul datelor generate în sistemele de sănătate este foarte mare. Dar există o problemă presantă cu generare de "date murdare", care cresc odată cu creșterea volumului de date, mai ales că cele mai multe sunt nestructurate și greu de utilizat. Utilizarea Big Data în domeniul asistenței medicale a generat provocări etice semnificative, cu implicații asupra drepturilor individuale, viața privată și autonomia, transparența și încrederea.

În media şi publicitate, pentru Big Data se folosesc numeroase puncte de informare despre milioane de persoane, pentru a servi sau transmite mesaje sau conținuturi personalizate.

În domeniul asigurărilor de sănătate se colectează date despre "factorii determinanți ai sănătății", care ajută la elaborarea de previziuni privind costurile de sănătate și identificarea problemele de sănătate ale clienților. Această utilizare este controversată, datorită discriminării clienților cu probleme de sănătate. (Allen 2018)

Big Data și tehnologia informației se complementează reciproc, ajutând împreună la dezvoltarea Internetului Lucrurilor (Internet of Things, IoT) pentru interconectarea dispozitivelor inteligente și colectarea datelor senzoriale utilizate în diferite domenii.

În sport, Big Data poate ajuta la îmbunătăți pregătirii și înțelegerea concurenților utilizând senzori specifici, și se poate prezice performanța viitoare a sportivilor. Senzorii atașați mașinilor din Formula 1 colectează, printre altele, date din presiunea în anvelope pentru a eficientiza arderea combustibilului.

2.1.1 În cercetare

În știință, sistemele Big Data sunt folosite intens în acceleratoarele de particule de la CERN (150 de milioane de senzori transmit date de 40 de milioane de ori pe secundă, pentru cca 600 de milioane de coliziuni pe secundă, din care se utilizează după filtrare doar 0,001% din totalul datelor obținute), (Brumfiel 2011) în telescoapele radio astrofizice construite din mii de antene, decodificarea genomului uman (inițial a durat câţiva ani, cu Big Data se poate realiza în mai puțin de o zi), studii climatice, etc.

Marile firme IT utilizează depozite de date de ordinul zecilor de petabyte pentru căutare, recomandări și merchandising. Cele mai multe date sunt colectate de Facebook,

cu peste 2 miliarde de utilizatori activi lunar, (Constine 2017) şi Google cu peste 100 de miliarde de căutări pe lună. (Sullivan 2015)

În cercetare se folosește mult căutarea criptată și formarea clusterelor în Big Data. Țările dezvoltate investesc enorm în prezent pentru cercetare în Big Data. În cadrul Uniunii Europene, aceste cercetări sunt înglobate în programul-cadrul Orizont 2020. (European Commission 2019)

Adesea, programele de cercetare folosesc resursele API de la Google şi Twitter pentru a obține acces la sistemele lor Big Data, gratuit sau contra cost.

Seturile mari de date vin cu provocări algoritmice care anterior nu existau, fiind imperios necesar să se schimbe în mod fundamental modalitățile de procesare. Pentru aceasta s-au creat ateliere speciale de lucru care reunesc oameni de știință, statisticieni, matematicieni şi practicieni pentru a discuta despre provocările algoritmice ale Big Data.

3. Aspecte filosofice

Big Data poate genera, prin inferențe, noi cunoașteri și perspective. Paradigma care rezultă din utilizarea Big Data generează noi oportunități.

Un motiv de îngrijorare majoră în cazul Big Data se datorează faptului că oamenii de știință de date tind să lucreze cu date despre subiectele pe care nu le cunosc şi cu care nu au fost niciodată în contact, fiind înstrăinați de produsul final al activității lor (aplicarea analizelor). Un studiu recent (Tanner 2014) afirmă că ceasta poate fi motivul unui fenomen cunoscut ca alienarea digitală.

Big Data are influențe mari la nivel guvernamental, afectând pozitiv societatea. Aceste sisteme pot fi eficientizate prin aplicarea politicilor de transparență și de guvernare deschisă, precum Open Data.

După elaborarea modelelor predictive privind comportamentul audienței țintă, Big Data pot fi utilizate pentru a genera avertizări timpurii pentru diverse situații. Există astfel un feedback pozitiv între cercetare și practică, cu descoperiri rapid preluate de practică.

A. Richterich afirmă că popularizarea monitorizării activității utilizatorilor a fost motivată de pretențiile că utilizarea (și colectarea de date cu) aceste dispozitive ar îmbunătăți bunăstarea, sănătatea și speranța de viață a utilizatorilor, și ar reduce semnificativ costurile asistenței medicale. (Richterich, 2018) Pentru a obține consimțământul utilizatorilor, multe firme au oferit reduceri acelor clienți care ar fi dispuși să furnizeze acces la datele lor de monitorizare. (Mearian 2015) Dar există și îngrijorări cu privire la influența acestor tehnologii asupra societății în special în probleme legate de corectitudine, discriminare, intimitate, abuz de date și siguranță. (Collins 2016)

Conceptual, Big Data ar trebui înțelese ca un termen umbrelă pentru un set de tehnologii emergente. În utilizarea lor, trebuie să ținem seama de contextele, rețelele, infrastructurile și interdependențele culturale, sociale și tehnologice care pot avea sens asupra Big Data. Termenul de "Big Data" nu se referă doar la datele ca atare, ci și la practicile, infrastructurile, rețelele și politicile care influențează diversele lor manifestări. Înțelegerea datelor mari ca un set de tehnologii emergente pare a fi utilă din punct de vedere conceptual, deoarece cuprinde evoluții digitale activate în colectarea, analiza și utilizarea datelor.” (Richterich, 2018)

În acest context, Rip descrie dilema evoluțiilor tehnologice: "Pentru tehnologiile emergente cu viitorul lor nedeterminat, există provocarea de a articula valori și reguli corespunzătoare care vor purta greutate. Acest lucru se întâmplă prin articularea

promisiunilor și viziunilor despre noile tehnologii [...] " (Rip 2013, 192) Astfel, tehnologiile emergente sunt locuri de" normativitate omniprezentă "caracterizată prin articularea promisiunilor și a temerilor, conceptualizând o astfel de "normativitate omniprezentă" ca o abordare "în spiritul eticii pragmatice, în care pozițiile normative co-evoluează." (Rip 2013, 205)

Etica pragmatică subliniază faptul că noile tehnologii se dezvoltă în societățile în care sunt discursiv asociate/disociate de anumite norme și valori. În același timp, pragmatismul afirmă că creșterea numărului mare de date și a practicilor legate de cercetare nu este o simplă chestiune a superiorității tehnologice. Ele formează un câmp de justificare normativă și contestație.

Neo-pragmaticii în abordarea eticii abordează cunoașterea epistemologică prin prisma falsificabilității cunoștințelor (științifice), cu evaluări critice ale structurilor puterii sociale. Keulartz et al. au propus o abordare pragmatică a eticii într-o cultură tehnologică (Keulartz et al. 2004) "ca o alternativă care combină punctele forte ale eticii aplicate și studiile științifice și tehnologice, evitând în același timp deficiențele acestor domenii." (Richterich, 2018) Astfel, etica aplicată este o abordare eficientă în ceea ce privește detectarea și exprimarea normativelor implicate în (inter-)acțiunile socio-tehnice sau care rezultă din acțiuni socio-tehnice, dar nu are posibilități în a surprinde normativitatea inerentă și agentul tehnologiilor. (Keulartz et al. 2004, 5)

Keulartz et al. consideră că astfel se poate depăși lipsa evaluărilor tehnologice normative: "impasul care a apărut din acest punct de vedere" (adică respectivele "pete oarbe" ale eticii aplicate) poate fi depășit printr-o reevaluare a pragmatismului." (Keulartz et al. 2004, 14) Pragmatismul etic poate fi caracterizat prin trei principii și principii comune: anti-fundaționalism, anti-dualism și anti-scepticism.

Anti-fundaționalismul se referă la principiul falsificabilității, considerând că nu putem ajunge la certitudine în ceea ce privește cunoștințele sau valorile ("adevăr final"), dar cunoașterea, ca și valorile și normele, se schimbă în timp. Valorile morale nu sunt statice, ci pot fi renegociate în funcție de evoluțiile tehnologice.

Anti-dualismul implică necesitatea de a se abține de la dihotomii predefinite. Printre dualismele criticate de Keulartz sunt esența/aparența, teoria/practica, conștiința/realitatea și faptele/valoarea. Etica aplicată tinde să asume astfel de dualisme ca a priori, spre deosebire de pragmatism care subliniază interrelațiile și liniile neclare dintre astfel de categorii.

Anti-scepticismul este strâns legat de nevoia de perspective situate și normativitate explicită, având legătură cu fundația anti-carteziană a pragmatismului.

În cercetarea europeană, pragmatismul a fost respins de obicei ca "superficial și oportunist", fiind asociat cu "stereotipuri negative", (Joas 1993) fiind acuzat de "utilitarism și meliorism." (Keulartz et al. 2004, 15) La sfârșitul anilor 1990 și 2000, pragmatismul a cunoscut o revigorare în cercetarea europeană. (Baert and Turner 2004)

Analiza Big Data din punct de vedere etic implică două aspecte principale interdependente: unul teoretic (descrierea filosofică a elementelor supuse controlului etic) și o viziune pragmatică (a impactului asupra vieții oamenilor și organizațiilor). (European Economic and Social Committee 2017)

Există probleme etice provocate de inteligența artificială, și o strânsă legătură între Big Data și inteligența artificială și derivatele acesteia: învățarea automată, analiza semantică, exploatarea datelor.

O abordare a eticii este prin agenția morală cu cel puțin cele trei condiții de cauzalitate, cunoaștere și alegere. Conform lui Noorman: (Noorman 2012)

- Există legături cauzale între persoane și rezultatul acțiunilor. Responsabilitatea persoanei derivă din controlul asupra rezultatului.
- Subiectul trebuie să fie informat, inclusiv despre posibilele consecințe.
- Subiectul trebuie să își dea consimțământul și să acționeze într-un anumit mod.

Profesorul Floridi, în *A patra revoluție*, identifică problema morală a Big Data cu descoperirea unui model simplu: o nouă frontieră a inovării și a concurenței. (Floridi 2014) O altă problemă asociată cu Big Data este riscul de a se descoperi aceste tipare, modificând astfel predicțiile.

Regula de bază a eticii Big Data este protecția vieții private, libertatea și puterea discreționară de a decide în mod autonom. De remarcat că există o tensiune continuă între nevoile individuale și cele ale unei comunități.

Este posibil să se identifice mai multe probleme etice care derivă din exploatarea Big Data: (European Economic and Social Committee 2017)

- *Confidențialitatea* - Limita extremă a confidențialității este seclusia, definită de Alan F. Westin ca "retragerea voluntară a unei persoane din societatea generală prin mijloace fizice într-o stare de singurătate". Moor și Tavani au definit un model de confidențialitate numit Control acces restricționat (RALC) care face diferența între confidențialitate, justificare, și managementul vieții private.
- *Realitatea adaptată și baloane de filtru* - Aplicația aflată pe un server colectează informații învățând din acestea, și apoi folosește acele informații pentru a construi un model al intereselor noastre. Când un sistem utilizează aceste modele pentru a filtra informații, e posibil să fim induși să credem că ceea ce vedem este o viziune completă a unui context specific, când de fapt noi suntem limitați de "înțelegerea" unui algoritm care a construit modelul. Efectele etice pot fi multiple: se pot

ascunde unele informaţii impunând prejudecăţi de care nu ştim, viziunea noastră asupra lumii poate deveni progresiv limitată, iar pe termen lung ar putea genera o favorizare a unui anumit punct de vedere.

- *Gestionarea ulterioară a datelor după deces* - Ce se întâmplă cu datele unui utilizator care a decedat? Moştenitorii devin proprietarii acestora? Se pot elimina datele din lumea digitală? Există aici probleme atât juridice cât și tehnologice.
- *Prejudecăți de algoritm* - Interpretarea datelor implică, aproape întotdeauna, anumite prejudecăţi. În plus, există posibilitatea ca o eroare într-un algoritm să introducă forme de părtinire. O problemă etică este încrederea noastră implicită în algoritmi, cu riscuri mari atunci când nu se iau în considerare riscurile datorită erorilor de programare sau rulare a algoritmilor.
- *Confidențialitatea vs. creşterea puterii de analiză* - Se referă la natura emergentă a informațiilor ca la un sistem complex: rezultatul datelor provenite din contexte diferite este mai mult decât suma simplă a părților.
- *Limitarea scopului* - Este foarte greu sau chiar imposibil să se limiteze utilizarea datelor. Confidenţialitatea nu este un singur element bloc, existând forme subtile de pierdere a vieţii private.
- *Inerția profilului digital al utilizatorilor* - Aceasta ține de subiectul realității personalizate. Un model care implică interesele unui utilizator se bazează, de obicei, pe comportamentul trecut și pe informațiile din trecut. Astfel, algoritmii nu se bazează pe identitatea actuală a persoanei, ci pe o versiune anterioară. Astfel se va influența comportamentul real utilizatorului, fiind împinși să își mențină interesele vechi și deci să nu poată descoperi alte oportunități. Dacă utilizatorul nu este conștient de această problemă, influența inerției va fi mult mai mare.

- *Radicalizarea utilizatorilor, conformismul şi sectarismul* - Big Data poate forma opinii folosind algoritmi de filtrare/recomandare, informaţii, articole și postări personalizate, și recomandări specifice de prieteni. Astfel, utilizatorii vor fi din ce în ce mai mult în contact cu oamenii, opiniile și faptele care vor susține poziția lor inițială. Această tendință este, de multe ori, ascunsă utilizatorilor sistemelor bazate pe Big Data, existând tendința dezvoltării unor prejudecăți, mergând de la conformism până la radicalizare. Se poate postula formarea unui fel de subconștient tehnologic cu impact asupra dezvoltării personalității utilizatorilor, fenomene evidente în cazul reţelelor sociale, unde distanţa dintre lumea reală ("fizică") și Internet este puternic atenuată.
- *Impact asupra capacităților personale și a libertății*
- *Drepturi egale între proprietarul datelor și exploatatorul de date* - De obicei persoana ale cărei date sunt utilizate nu este proprietarul legal al acestora. De aceea, o cerință minimă este ca acea persoană să aibă acces la propriile date, permițându-i să le descarce și eventual să le șteargă.

4 Aspecte legale

Utilizarea Big Data prezintă probleme juridice semnificative, în special din punctul de vedere al protecției datelor. Cadrul juridic existent al Uniunii Europene, bazat în special pe Directiva nr. 46/95/CE și Regulamentul general privind protecția datelor cu caracter personal, oferă o protecție corespunzătoare. Dar, pentru Big Data este necesară o strategie cuprinzătoare şi globală. Evoluția în timp a fost de la dreptul de a exclude pe alții la dreptul la controlul propriilor date și, în prezent, la regândirea dreptului la identitate (digitală).

Colectarea și agregarea datelor în Big Data nu sunt supuse reglementărilor privind protecția datelor, datorită noilor perspective privind confidențialitatea, cu posibilitatea apariției unor forme specifice de discriminare.

În 2014, în raportul Podesta s-a concluzionat că "analizele Big Data au potențialul de a eclipsa protecția drepturilor pe termen lung în ceea ce privește modul în care informațiile personale sunt utilizate în locuințe, pentru credite, ocuparea forței de muncă, sănătate, educație." (European Economic and Social Committee 2017) Rezultă că sunt necesare noi modalități specifice de protecție a cetățenilor, deoarece cadrul juridic, chiar dacă este teoretic aplicabil, nu pare să ofere o protecție adecvată și deplină.

4.1 GDPR

Regulamentul UE privind protecția datelor 2016/679 (General Data Protection Regulation, "GDPR") se ocupă de protecția datelor și viața privată a persoanelor din Uniunea Europeană și Spațiul Economic European. Abordează în mod special exportul de date cu caracter personal în afara zonelor UE și SEE. GDPR intenționează să simplifice mediul de reglementare prin unificarea reglementării în cadrul UE. (European Parliament 2016)

GDPR se aplică în două cazuri pentru prelucrarea datelor cu caracter personal (a) accesare de bunuri sau servicii contra cost de persoane din UE, sau (b) monitorizarea comportamentului lor în cadrul UE. Astfel, regulamentul permite extinderea acestuia la toți furnizorii de servicii pe internet, chiar dacă nu sunt stabiliți în UE. Mai general, GDPR se aplică tuturor agregatorilor mari de date, indiferent de conexiunile geografice sau fizice.

Etapele procesării datelor personale

Prelucrarea datelor cu caracter personal este definită de articolul 4, alineatul (2), ca fiind "orice operațiune sau set de operațiuni care se efectuează pe date cu caracter personal sau pe seturi de date cu caracter personal, cum ar fi colectarea, înregistrarea, organizarea, structurarea, stocarea, adaptarea sau modificarea, recuperarea, consultarea, utilizarea, dezvăluirea prin transmitere, difuzarea sau punerea la dispoziție în alt mod, alinierea sau combinarea, restricționarea, ștergerea sau distrugerea acestor date".

Big Data include mai multe activități de prelucrare a datelor cu caracter personal, fiecare cu normele ei specifice:

1. colectarea datelor
2. stocarea datelor
3. agregarea datelor
4. analiza datelor și utilizarea rezultatelor analizei

Principiile procesării datelor

Prelucrarea datelor se bazează pe următoarele principii stabilite în Articolul 5 din GDPR:

1. Legalitate, corectitudine și transparență: Utilizatorii trebuie să fie pe deplin și corect informați în ceea ce privește politica de confidențialitate și, să poată accesa ușor propriile date.
2. Limitarea scopului: Colectorii de date trebuie să informeze persoana vizată despre scopurile colectării datelor, care pot fi prelucrate ulterior numai în acele scopuri.
3. Minimizarea datelor: Vor fi colectate numai datele cu caracter personal relevante scopurilor declarate.

4. Precizia și actualizarea: Datele se vor actualiza și rectifica ori de câte ori se impune prin scopul declarat. În cazul Big Data, este foarte important dreptul utilizatorilor de anulare sau ștergere a datelor personale.
5. Limitarea stocării: Datele vor fi păstrate doar pe perioada procesării și ulterior vor fi distruse. Durata stocării poate fi prelungită în măsura în care datele sunt arhivate în scopuri de interes public, cercetare științifică sau istorică sau statistice.
6. Integritatea și confidențialitatea: operatorul de date: Se va asigura o securitate adecvată datelor cu caracter personal prin măsuri tehnice și organizatorice.

Politica de confidențialitate și transparența

În cazul colectării de date pentru a completa un formular, se va respecta principiul minimizării datelor, solicitându-se doar datele relevante și strict necesare. În cazul colectării automate de date, precum cookie, monitorizare web sau geolocație, politica de confidențialitate trebuie să informeze utilizatorul asupra acestui aspect.

Scopurile procesării datelor

Datele anonime și agregate pot fi procesate în scopul identificării comportamentului anumitor categorii de consumatori. În acest scop, operatorul de date efectuează anonimizarea și le transferă apoi unei terțe părți care le utilizează.

Confidențialitate prin design și implicită

Conceptele de confidențialitate prin design și confidențialitate implicită nu au fost incluse explicit în reglementările UE. Dar, în conformitate cu art. 78 din GDPR, "Pentru a putea demonstra conformitatea cu prezentul regulament, operatorul ar trebui să adopte politici interne și să pună în aplicare măsuri care să îndeplinească, în special, principiile protecției datelor prin proiectare și protecția datelor în mod implicit. Astfel de măsuri ar

putea consta, între altele, în minimizarea prelucrării datelor cu caracter personal, pseudonimizarea datelor cu caracter personal cât mai curând posibil, transparența în ceea ce privește funcțiile și prelucrarea datelor cu caracter personal, care să permită persoanei vizate să monitorizeze prelucrarea datelor, permițând operatorului să creeze și să îmbunătățească caracteristicile de securitate. La elaborarea, proiectarea, selectarea și utilizarea aplicațiilor, serviciilor și produselor care se bazează pe prelucrarea datelor cu caracter personal sau prelucrarea datelor cu caracter personal pentru a-și îndeplini sarcinile, ofertanții de produse, servicii și aplicații ar trebui încurajați să ia în considerare dreptul la date atunci când proiectează și elaborează astfel de produse, servicii și aplicații, ținând cont de stadiul actual al tehnicii, pentru a se asigura că operatorii și prelucrătorii își pot îndeplini obligațiile de protecție a datelor."

Paradoxul (legal) al Big Data

Utilizarea Big Data implică cel puțin un paradox: pe de o parte, Big Data asigură o transparență maximă dar, în același timp, nu există o transparență adecvată în ceea ce privește utilizarea Big Data. Transparența este o problemă fundamentală deoarece influențează capacitatea unui utilizator de a permite divulgarea informațiilor sale.

5. Probleme etice

Etica Big Data presupune aderarea la conceptele de comportament corect și greșit în ceea ce privește datele, în special datele cu caracter personal. Etica Big Data pune accentul pe colectorii și diseminatorii de date structurate sau nestructurate.

Etica Big Data este susținută, la nivelul UE, de o amplă documentație, prin care se încearcă să se găsească soluții concrete pentru maximizarea valorii Big Data fără a sacrifica drepturile fundamentale ale omului. Autoritatea Europeană pentru Protecția

Datelor (AEPD) sprijină dreptul la viață privată și dreptul la protecția datelor cu caracter personal în respectul demnității umane. Conform acestor documente, trebuie să se depășească conflictul conceptual dintre viața privată și Big Data,și între intimitate și inovație. Este esențial să se identifice modalitățile de includere a dimensiunii etice în conceperea inovațiilor. (European Economic and Social Committee 2017)

Conform noului Regulament al UE 2016/679, operatorii de date trebuie să pună în aplicare măsurile de confidențialitate și tehnologiile de îmbunătățire a confidențialității în momentul determinării modalităților de procesare și al procesării în sine. Prin ENISA75 au fost identificate multe strategii de confidențialitate prin design (minimizarea datelor, ascunderea datelor cu caracter personal și interconexiunile acestora, prelucrarea separată a datelor cu caracter personal, alegerea celui mai înalt nivel de agregare, transparența, monitorizarea, politica de confidențialitate, aspecte legale).

O modalitate de bază pentru coexistența pașnică dintre exploatarea Big Data și protecția datelor este *controlul*, de către utilizator, a datelor personale, ceea ce conduce la transparență și încredere între utilizatori și furnizorii de servicii digitale. După cum s-a subliniat în evaluarea impactului GDPR,

"Construirea încrederii în mediul online este esențială pentru dezvoltarea economică. Lipsa de încredere îi face pe consumatori să ezite să cumpere online și să adopte noi servicii, inclusiv serviciile publice de e-guvernare. Dacă nu este abordată, această lipsă de încredere va continua să încetinească dezvoltarea utilizărilor inovatoare ale noilor tehnologii, să acționeze ca un obstacol în calea creșterii economice și să blocheze sectorul public în a profita de beneficiile potențiale ale digitalizării serviciilor sale."

În cazul Big Data modelele tradiționale de *consimțământ* sunt insuficiente și depășite. ”Consimțământul ar trebui să fie suficient de granular pentru a acoperi toate procesele diferite și scopurile de prelucrare și reutilizare a datelor cu caracter personal.” (European Economic and Social Committee 2017)

O problemă specială este *portabilitatea* datelor, susținută la nivelul UE de AEPD în Avizul 7/2015, (MORO 2016) unde se impune garantarea dreptului cetățenilor de a accesa și corecta datele personale printr-un control extins. Portabilitatea datelor poate ajuta la creșterea gradului de conștientizare și control al consumatorilor prin transferul între servicii online.

AEPD consideră că datele cu caracter personal ar trebui să fie tratate la fel ca alte resurse importante, precum petrolul, unde tranzacționarea are loc între părți la fel de bine informate (simetria informațională). În realitate, piața informațiilor cu caracter personal are un caracter de asimetrie informațională, nefiind nici transparentă, nici echitabilă, clienții nefiind compensați pentru informațiile personale pe care le oferă. Astfel, portabilitatea datelor ar încuraja un mediu mai competitiv între beneficiarii acestor date, utilizatorii având posibilitatea să aleagă cui oferă atele personale.

O altă abordare pune în discuție *stocarea* datelor cu caracter personal, cu posibilitatea pentru utilizator de a acorda sau retrage consimțământul pentru datele sale personale. (MORO 2016) (DG Connect 2015) Stocarea datelor cu caracter personal implică un ”concept cadru și o implementare arhitecturală care transferă achiziția și controlul datelor de la un model de date distribuit la un model orientat spre utilizator.” (European Economic and Social Committee 2017) Portabilitatea datelor ar putea asigura acest deziderat.

AEPD susține promovarea beneficiarilor responsabili și reducerea birocrației în protecția datelor, prin coduri de conduită, audituri, certificări, și o nouă generație de clauze contractuale și reguli corporative obligatorii. Responsabilitatea beneficiarilor Big Data presupune instituirea unor politici interne și a unor sisteme de control conforme cu legislația în vigoare, prin soluții inteligente și dinamice care să garanteze respectarea

principiilor fundamentale (minimizarea datelor, limitarea scopului, calitatea datelor, procesarea corectă și transparentă a datelor, design, limitare de stocare, integritate și confidențialitate).

Etica datelor se bazează pe următoarele principii: **proprietatea** (persoanele fizice dețin propriile date**, transparența tranzacțiilor** (utilizatorii trebuie să aibă acces transparent la proiectarea algoritmului**), consimțământ** (utilizatorul trebuie să fie informat și să își exprime explicit consimțământul cu privire la utilizarea datelor personale**, confidențialitate** (trebuie protejată confidențialitatea utilizatorilor)**, financiar** (utilizatorul să cunoască tranzacțiile financiare rezultate din utilizarea datelor lui personale**), și deschidere** (seturile de date agregate să fie disponibile în mod liber).

Etica în cercetare

Termenul de studiu critic de date (SCD) implică faptul că cercetătorii investighează Big Data din perspective critice. Studierea datelor în acest context implică, pe lângă analiza lor, și încorporarea datelor în practici (cunoașterea), instituții și sisteme politice și economice, prin interacțiunea complexă dintre date și entitățile care le produc, dețin și folosesc.

Un raport al OECD (2013) subliniază că, spre deosebire de normele etice aplicate datelor obișnuite de cercetare, în cazul Big Data: (OECD 2013)

- Colectarea de date nu a făcut obiectul unui proces formal de examinare etică.
- Normele etice obișnuite nu vor fi implementate în cazul Big Data
- Utilizarea datelor pentru cercetare poate să difere de scopul inițial.
- Datele nu mai sunt deținute ca seturi discrete.

Relația dintre cei care oferă datele și cei care le folosesc este adesea indirectă și variabilă. Un raport mai recent al OECD (2016) susține că această relație este mai slabă sau inexistentă, Big Data limitând capabilitățile obișnuite. (OECD 2016)

Stocarea datelor e importantă pentru integritatea cercetării. Datele trebuie să aibă o "proveniență" clară, cu surse și procesare cunoscute, identificate și documentate.

Multe date care nu sunt colectate special pentru cercetare au standarde diferite în cercetarea datelor.

Pentru anumite date, adesea cu valoare comercială (de ex., datele colectate pe Twitter), există restricții legale privind reproducerea lor. (UK Data Service 2017)

Depozitele de date trebuie să respecte standardele de transparență și reproductibilitate.

Conștientizarea

Conștientizarea tipului de date care sunt furnizate în timpul unei înregistrări online (pentru crearea unui cont, sau un abonament, de ex.) este un fapt rar, mai ales că există posibilitatea folosirii unei identități digitale deja existente (profil Facebook, de ex.) în locul unei înregistrări separate, pentru un acces mai rapid. Astfel de situații creează o opacitate cu privire la datele partajate între furnizorul de identitate și serviciul utilizat.

Consimțământul

Pentru utilizarea datelor cu caracter personal ale unei persoane, este nevoie de consimțământul informat și explicit exprimat al acesteia referitor la cine, când, cum și în ce scop se folosesc. Când trebuie partajate datele, aceste utilizări trebuie aduse la cunoștința persoanei. Ar trebui să fie întotdeauna posibilă retragerea consimțământului pentru viitoarele utilizări.

În analizele Big Data, se poate cunoaşte foarte puţin despre utilizările viitoare intenţionate ale datelor, și despre beneficiile, și riscurile implicate. Aici, există proceduri pentru consimțământul "larg" și "generic" de a împărtăşi datele genomice, de ex., și în scopuri diferite. Chiar şi atunci când se procedează corect, există anumite provocări practice specifice: obținerea consimțământului în cunoștință de cauză poate fi imposibil sau foarte costisitor, iar valabilitatea consimțământului este disputabil când acordul este obligatoriu pentru a accesa un serviciu.

Controlul

În lumea actuală, datele personale pot fi tranzacţionate la fel ca orice monedă în implementarea Big Data. Există opinii diferite în ce măsură această situație este una etică, inclusiv cine să participe la profitul obținut din aceste tranzacţionări..

În modelul de tranzacţionare a datelor cu caracter personal, transmiterea datelor personale este un cadru care oferă persoanelor posibilitatea de a-și controla identitatea digitală și a crea acorduri granulare de partajare a datelor.

În prezent prinde contur ideea datelor deschise, centrată în jurul argumentului că datele ar trebui să fie disponibile în mod liber. Dorința de a partaja date variază în funcție de persoană.

În cazul copiilor, părinții sau tutorii au responsabilitatea pentru datele lor, care nu pot fi tranzacționate contra beneficii financiare.

La nivel naţional, un guvern este suveran asupra datelor generate și colectate. La 26 octombrie 2001 a intrat în vigoare Actul Patriot în SUA, iar la 25 mai 2018, Regulamentul general privind protecţia datelor 2016/679 (GDPR) la nivelul Uniunii Europene, pentru problemele legate de protecția datelor personale.

În Big Data, relația om-date este asimetrică, bazată pe controlul datelor. ”Dreptul de a fi uitat”, adoptat la nivelul UE, este unul din elementele de bază ale controlului unui individ asupra datelor sale personale.

Transparența

Algoritmii utilizați în Big Data pot determina prejudecăţi care afectează sistematic drepturile individului. De aceea, proiectarea algoritmului ar trebui să fie transparentă și inclusivă.

Guvernarea anticipativă implică analize predictive pe baza Big Data pentru a evalua potenţiale comportamente, cu implicaţii etice care pot încuraja prejudecăţile şi discriminarea.

O persoană care acceptă includerea datelor sale personale în Big Data are dreptul să ştie de ce se colectează datele, cum vor fi folosite, cât timp vor fi stocate, și cum pot fi modificate.

Încrederea

Încrederea în sistemele Big Data este legată de interdependenţe cu confidențialitatea și conștientizarea. Până în prezent, încrederea a fost considerată din perspectivă strict tehnologică. Se speră să se realizeze arhitecturi hardware și software care ar putea creşte încredere între fiinţe umane și obiecte, și deci o mai mare acceptanţă a utilizării datelor personale.

Proprietatea

O întrebare fundamentală în etica cercetării cu Big Data este, cine deține datele? Aceasta implică subiectul drepturilor şi obligaţiilor asupra proprietăţii. În legislaţia europeană, GDPR indică faptul că persoanele dețin propriile date cu caracter personal.

Suma datelor personale ale unui individ formează o identitate digitală.

Protecţia drepturilor morale (dreptul de a fi identificat ca sursă a datelor, şi de a le controla) ale unui individ se bazează pe opinia că datele personale sunt o expresie directă a personalităţii acestuia, şi nu pot fi transferate unei alte persoane decât, eventual, prin succesiune atunci când individul moare.

Proprietatea implică exclusivitate, respectiv restricţionarea implicită a altora în ceea ce privește accesul la proprietate. O proprietate eficientă a datelor personale implică portabilitatea, posibilitatea de a folosi alternative fără a pierde din date. Standardizarea ar ajuta, de asemenea, la curățare datelor personale.

În mod efectiv, în prezent, datele sunt deținute de proprietarul senzorilor, cel care efectuează înregistrarea sau entitatea care deține senzorul.

În UE, s-a restrâns în mod progresiv posibilitatea ca datele cetățenilor UE să fie stocate în afara așa-numitului "Euro cloud", dar nu s-a rezolvat problema datelor deja stocate și prelucrate în altă parte, și ”nu rezolvă dilema etică a modului în care proprietatea asupra datelor este definită în mod filosofic, înainte de a trece la o abordare mai degrabă a legii și a elaborării politicilor.” (European Economic and Social Committee 2017)

Supravegherea și securitatea

Din ce în ce mai multe surse de date sunt disponibile cu ajutorul tehnologiilor avansate, precum circuitele CCTV, GPS, dispozitive mobile, carduri de credit, ATM. De asemenea, supravegherea activă este o metodă de colectare a datelor, dar în același timp de limitare a libertăților cetățenilor. O astfel de supraveghere permanentă determină

creșterea stresului oamenilor, și creează tendința acestora de a se comporta într-un anumit mod care să se conformeze normelor așteptate.

Identitatea digitală

Identitatea digitală are avantajul accesului rapid la conținutul online și serviciile conexe. Utilizarea identității digitale are potențialul de a genera discriminare bazată pe reprezentarea unei persoane conform datelor ei online, care de multe ori poate să nu corespundă cu situația reală, într-un proces numit ”dictatura datelor” în care "nu mai suntem judecați pe baza acțiunilor noastre, ci pe baza a ceea ce indică toate datele despre noi ca fiind acțiunile noastre probabile", (Norwegian Data Protection Authority 2013) interacțiunea personală nefiind plasată într-un plan secundar.

Realitatea ajustată

Orice interacțiune a noastră cu Internetul implică posibilitatea stocării datelor noastre personale. Prelucrarea și analiza acestor date determină rezultatele personalizate care ne apar ulterior pe Internet, prin rezultate ale căutărilor noastre, afișarea produselor în magazinele online, afișarea reclamelor, etc. Se generează astfel o versiune mai îngustă și mai personalizată a experienței online anterioare a unui utilizator (așa-numitul "balon de filtrare" (Pariser 2011)). Un avantaj este că utilizatorul va găsi rapid ceea ce caută de obicei, dar excluderea anumitor aspecte, perspective și idei poate duce la o restrângere a creativității și dezvoltarea unei atitudini tolerante prin izolarea politică și socială de celelalte aspecte, prin lipsa unor viziuni pluraliste. (Crawford, Gray, and Miltner 2014)

De-anonimizarea

De-identificarea implică ștergerea sau ascunderea elementelor care ar putea identifica imediat o persoană sau organizație. Legislația din diferite țări privind protecția

datelor defineşte tratamente diferite pentru datele identificabile. Identificabilitatea este văzută din ce în ce mai mult ca un continuum, nu un aspect binar. Riscurile de divulgare cresc simultan cu numărul de variabile, de surse de date și cu puterea analizei datelor. Riscurile de dezvăluire pot fi atenuate, dar nu eliminate. De-identificarea rămâne un instrument vital pentru asigurarea utilizării în siguranță a datelor. (UK Data Service 2017)

Informații perfect anonime luate separat, pot fi combinate cu alte date pentru a identifica în mod unic o persoană cu grade diferite de certitudine. Profilarea poate deveni un instrument puternic, ridicând îngrijorări cu privire la gradul în care este permisă intruziunea în viața unui individ, posibilitatea asigurării securității, și supravegherea.

Inegalitatea digitală

Avantajele dimensiunii mari a datelor sunt clare, dar există și opinii conform cărora acumularea de date la o scară uriașă prezintă riscuri specifice. Din această cauză, sunt puține entități care au acces, prin infrastructură și abilități, la sistemele Big Data. În acest context, costurile și abilitățile necesare accesului duce la anumite inegalități digitale specifice abordate de etică.

Confidențialitatea

În tranzacțiile de date este foarte important să se asigure confidențialitatea:

"Nimeni nu va fi supus la interferenţe arbitrare cu intimitatea, familia, casa sau corespondența sa și nici la atacuri asupra onoarei și reputației sale. Toată lumea are dreptul la protecția legii împotriva unor asemenea ingerinţe sau atacuri. "- Declarația Organizației Națiunilor Unite privind Drepturile Omului, Articolul 12.

În multe țări, monitorizarea publică datelor de către guvern pentru a observa cetățenii necesită o autorizare explicită printr-un proces judiciar adecvat. Confidențialitatea nu este despre păstrarea secretelor, ci despre alegere, drepturile omului, și libertate.

Adesea confidenţialitatea este văzută în mod greşit ca o alegere binară între izolare şi progres ştiinţific. Protejarea identităţii în date este posibilă tehnologic, de exemplu utilizând criptarea homomorfă și designul algoritmic.

Confidenţialitatea ca o limitare a utilizării datelor poate fi, de asemenea, considerată ne-etică, (Kostkova et al. 2016) în special în asistenţa medicală, dar trebui ţinut cont de faptul că este posibilă extragerea valorii datelor fără a compromite intimitatea.

Confidențialitatea este recunoscută ca un drept uman prin numeroase reglementări naţionale şi internaţionale. Confidenţialitatea în cercetare se realizează printr-o combinație de abordări: limitarea datelor colectate, anonimizarea acestora; și reglementarea accesului la date. În cazul cercetării Big Data apar probleme specifice: ambiguitatea între termenii "privațiune" și "confidențialitate; declararea spațiilor sociale ca publice sau private; necunoașterea riscurilor de confidențialitate de către utilizatori; distincția neclară între uzanțele publice și private. În prezent există dispute dacă știința datelor ar trebui să fie clasificată ca o cercetare a subiecților umani, și deci nesupusă normelor obișnuite de confidențialitate.

6. Cercetarea Big Data

Prin noile concepte de "daune algoritmice", "analize predictive", etc., algoritmii folosiți în prezent în operațiunile Big Data depășesc viziunea tradițională a confidențialității. Conform Consiliului Național pentru Știință și Tehnologie,

> ”"Algoritmii analitici" sunt algoritmi pentru prioritizare, clasificare, filtrare și predicție. Utilizarea acestora poate crea probleme de confidențialitate atunci când informațiile utilizate de algoritmi sunt inadecvate sau inexacte, atunci când apar decizii incorecte, atunci când nu există mijloace rezonabile de recurs, atunci când autonomia unui individ este direct legată de rezultatul algoritmic sau atunci când

> folosirea algoritmilor predictivi încurajează alte daune asupra vieții private.” (NSTC (National Science and Technology Council) 2016, 18)

Cercetările Big Data sunt ceea ce eticianul James Moor ar numi "harababura conceptuală" datorită ”incapacității de a conceptualiza în mod corect valorile etice și dilemele de joc într-un context tehnologic nou.” (Buchanan and Zimmer 2018) În această situație confidențialitatea este asigurată printr-o combinație de diferite tactici și practici (medii controlate sau anonime, limitarea informațiilor personale, anonimizarea datelor, restricții de acces, securizarea datelor, etc.). În general, toate noțiunile conexe devin confuze în cazul Big Data. Astfel, postările sociale sunt considerate publice în rețelele sociale în cazul unei setări corespunzătoare. Dar rețelele sociale sunt medii complexe de interacțiuni socio-tehnice unde utilizatorii nu înțeleg întotdeauna funcționalitatea setărilor și termenii de utilizare. Astfel, există o incertitudine în ceea ce privește intențiile și așteptările utilizatorilor, iar aceste deficiențe conceptuale în contextul cercetărilor Big Data conduc la incertitudini în ceea ce privește necesitatea consimțământului informat.

Concluzii

Studiile de date critice în Big Data reflectă practicile, culturile, politica și economiile specifice. (Dalton, Taylor, and Thatcher 2016) Problemele pot varia de la intimitatea și autonomia indivizilor la etica științei datelor și schimbările instituționale datorate cercetărilor Big Data. Rezultă necesitatea de a analiza practicile Big Data conștiente de relațiile de putere, prejudecăți și inegalități.

O definiție care ar restrânge cercetarea critică la domeniul teoriei normative și critice ar fi contraproductivă.

Principiile comune ale studiilor de date critice evidențiază interdependențele dintre tehnologiile emergente și actorii (umani) din societățile din ce în ce mai bine

prezentate. Big Data sunt și un produs al condițiilor socio-tehnice contemporane, deoarece sunt producătoare de astfel de condiții. (Richterich, 2018)

Domeniul studiilor de știință și tehnologie (STS) are o relație destul de ambiguă cu evaluările normative ale tehnologiei.

În STS, anumite componente sunt preocupate mai mult de abordările descriptive decât de cele normative.

Spre deosebire de idealul STS comun al unui relativism "fără valoare", (Pels 1996, 277) Pels solicită recunoașterea "pozițiilor a treia" în evaluările producției de cunoștințe științifice care ” [...] nu sunt exterioare domeniului de controversă studiat, ci sunt incluse și implicate în acesta. [...] Ele nu sunt libere de valoare sau dispărute, ci sunt situate, parțial și comise în sens politic și de cunoaștere." (Pels 1996)

O problemă majoră în Big Data este că nu se cunosc prea bine micro-procesele empirice care stau la baza apariției caracteristicilor lor tipice de rețea. (Snijders, Matzat, and Reips 2012) Big Data trebuie întotdeauna contextualizate în contextele sociale, economice și politice ale acestora. (Graham 2012)

Susținătorii protejării vieții private sunt îngrijorați de amenințarea la adresa vieții private datorită creșterii volumului de stocare și integrare a informațiilor de identificare personală. În acest sens, există diferite recomandări politice pentru a conforma practica și viața privată. (Ohm 2012) Folosirea abuzivă a Big Data de mass-media, companii și chiar guvern a dus la pierderea încrederii în instituțiile sociale. Pentru a proteja libertățile individuale, Nayef Al-Rodhan consideră că este necesar un nou tip de contract social, cu monitorizarea și reglementarea mai atentă a Big Data. (Al-Rodhan 2018)

Experimentele ştiinţifice au tendinţa de a analiza datele utilizând clustere specializate și computere de înaltă performanță, mai degrabă decât cloud, diferind astfel cultural și tehnologic de restul societății.

Utilizarea Big Data, datorită manipulării cantităților mari de date, a dus la neglijarea principiilor ştiinţei, precum alegerea eșantioanelor reprezentative, determinând prejudecăți în analiza rezultatelor. Această analiză este adesea superficială în comparaţie cu analiza seturilor mai mici de date. (Piatetsky 2014) Unele surse de date, precum Twitter, nu sunt reprezentative pentru populaţia totală. Ioannidis a susţinut că în folosirea Big Data, "cele mai multe rezultate publicate în cercetare sunt false" (Ioannidis 2005) întrucât probabilitatea ca un rezultat "semnificativ" să fie fals creşte rapid cu volumul datelor, dar sunt publicate doar rezultatele pozitive.

În utilizarea Big Data, UK Data Service evidenţiază mai multe probleme etice specifice: (UK Data Service 2017)

- Au apărut alternative la consimțământul individual informat, de exemplu "consimțământul social", mult mai permisive.
- A crescut necesitatea respectării sursei datelor şi, în general, "integritatea contextuală", în cazul reutilizării datelor.
- Etica de cercetare se bazează în principal pe ideea că entitatea cercetată este o persoană individuală, deci ar fi posibilă de-identificarea pentru protejare. În cazul considerării unui grup ca un întreg, protecția socială scade. În acest caz s-a propus ca datele să fie considerate ca "beneficii publice" sau de "interes public", dar aceasta nu rezolvă responsabilitatea utilizatorii datelor.

Matthew Zook et al. propune "zece reguli simple" etice în utilizarea Big Data în cercetare. (Zook et al. 2017) Primele cinci reguli privesc modul de reducere a șanselor de

vătămare care rezultă din practicile de cercetare, iar celelalte reguli se referă la cele mai bune practici.

1. **Datele sunt oameni și pot face rău**: majoritatea datelor reprezintă sau influențează oamenii. Începeți cu presupunerea că datele sunt persoane (până când se dovedește altfel) și ghidați-vă analiza pe această bază.
2. **Confidențialitatea este mai mult decât o valoare binară**: confidențialitatea depinde de natura datelor, contextul în care au fost create și obținute, și de așteptările și normele celor afectați. Ea se extinde la grupuri. Contextualizați datele pentru a anticipa încălcarea confidențialității și pentru a minimiza daunele.
3. **Evitați reidentificarea datelor dvs.**: de multe ori nu se reușește anonimizarea eficientă a datelor. Datele considerate a fi anonime sunt combinate cu alte variabile care pot duce la re-identificare. Identificați vectorii posibili de reidentificare și minimizați-i în rezultatele publicate.
4. **Practicați schimbul de date etice:** Pentru unele proiecte, precum în genetică, schimbul de date este o necesitate socială, dar rămân în continuare valabile consimțământul informat și de dreptul de retragere. Partajați datele respectând protocoalele de cercetare, dar țineți cont de daunele potențiale generate de datele colectate informal.
5. **Luați în considerare punctele tari și limitele datelor dvs.; mai mare nu înseamnă automat mai bine**: seturile de date trebuie fundamentate în contextul lor adecvat, ținând cont inclusiv de conflictele de interese. În achiziția datelor, este important să se înțeleagă sursa datelor, și să se respecte reglementările. În medii cu reglementare slabă, pot fi folosite normele etice.

Cercetătorii trebuie să fie sensibili la semnificațiile multiple potențiale ale datelor. Documentați proveniența și evoluția datelor.

6. **Dezbateți alegerile dure, etice**: trebuie să se evite lipsa soluțiilor clare și a protocoalelor. Astfel de dezbateri pot produce evaluări *inter pares* foarte utile. Se pot utiliza serviciile de consultare în domeniul eticii de cercetare din universități. Implicați-vă colegii și studenții în practica etică pentru o cercetare de mare amploare a Big Data.
7. **Elaborați un cod de conduită pentru organizația dvs., comunitatea de cercetare sau industria**: "etica falsă", ca și falsificarea datelor sau rezultatelor, sunt inacceptabile. Este necesară elaborarea de coduri de conduită, care pot oferi îndrumări în evaluarea reciprocă a publicațiilor și în examinarea finanțării. Stabiliți codurile adecvate de conduită etică, împreună cu reprezentanții comunităților afectate.
8. **Proiectați-vă datele și sistemele pentru audit**: auditul furnizează un mecanism de verificare a muncii, sporind înțelegerea și replicabilitatea. Planificați și inițiați audituri ale practicilor Big Data.
9. **Implicați-vă cu consecințe mai nari în practicile de date și analiză**: este important ca cercetătorii să gândească dincolo de valorile tradiționale. Se poate cere furnizorilor stocarea în cloud, și centrele de prelucrare a datelor pot să treacă la surse de energie durabile și regenerabile. Realizarea unor cercetări de mare amploare are efecte la nivel de societate.
10. **Să știți când să încălcați aceste reguli**: trebuie să știți la ce să vă așteptați când vă îndepărtați de aceste reguli, precum în situații de dezastru natural sau de

urgență. Cercetarea responsabilă a Big Data depinde de mai multe liste de verificare.

Indiferent de normele etice sau legale oamenii de știință trebuie să fie riguroși în utilizarea tehnicilor și metodologiilor, și foarte precauți în probleme de etică. Ideea că "datele sunt deja publice" (Zimmer 2016) sunt simplificări nejustificate. Datele nu sunt ceva abstract, ele sunt de fapt persoane reale.

Cercetarea responsabilă a Big Data nu vizează restrângerea cercetării, ci asigurarea încrederii, corectitudinii și maximizării aspectelor pozitive, reducând în același timp răul. Big Data oferă oportunități fantastice pentru a înțelege mai bine societatea și lumea, dar trebuie să se țină cont și de responsabilitatea eticii în alegerile, practicile și acțiunile cercetărilor.

Bibliografie

Allen, Marshall. 2018. "Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates." Text/html. ProPublica. July 17, 2018. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates.

Al-Rodhan, Nayef. 2018. "The Social Contract 2.0: Big Data and the Need to Guarantee Privacy." OpenMind. June 11, 2018. https://www.bbvaopenmind.com/en/humanities/beliefs/the-social-contract-2-0-big-data-and-the-need-to-guarantee-privacy-and-civil-liberties/.

Baert, Patrick, and Bryan Turner. 2004. "New Pragmatism and Old Europe: Introduction to the Debate between Pragmatist Philosophy and European Social and Political Theory." *European Journal of Social Theory* 7 (3): 267–74. https://doi.org/10.1177/1368431004044193.

Brumfiel, Geoff. 2011. "High-Energy Physics: Down the Petabyte Highway." *Nature* 469 (7330): 282–83. https://doi.org/10.1038/469282a.

Buchanan, Elizabeth A., and Michael Zimmer. 2018. "Internet Research Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/ethics-internet-research/.

Collins, Tom. 2016. "Security Fears Sparked over Wearable Technology." Mail Online. December 19, 2016. http://www.dailymail.co.uk/~/article-4049154/index.html.

Constine, Josh. 2017. "Facebook Now Has 2 Billion Monthly Users… and Responsibility." *TechCrunch* (blog). 2017. http://social.techcrunch.com/2017/06/27/facebook-2-billion-users/.

Crawford, Kate, Mary L. Gray, and Kate Miltner. 2014. "Big Data| Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction." *International Journal of Communication* 8 (0): 10. https://ijoc.org/index.php/ijoc/article/view/2167.

Dalton, Craig M., Linnet Taylor, and Jim Thatcher. 2016. "Critical Data Studies: A Dialog on Data and Space." In . https://doi.org/10.1177/2053951716648346.

Dean, Jeffrey, and Sanjay Ghemawat. 2004. "MapReduce: Simplified Data Processing on Large Clusters." http://static.googleusercontent.com/media/research.google.com/en//archive/mapreduce-osdi04.pdf.

Dedić, Nedim, and Clare Stanier. 2017. "Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery." In *Innovations in Enterprise Information Systems Management and Engineering*, edited by Felix Piazolo, Verena Geist, Lars Brehm, and Rainer Schmidt, 114–22. Lecture Notes in Business Information Processing. Springer International Publishing.

DG Connect. 2015. "Study on Personal Data Stores Conducted at the Cambridge University Judge Business School." Text. Digital Single Market - European Commission. August 7, 2015. https://ec.europa.eu/digital-single-market/en/news/study-personal-data-stores-conducted-cambridge-university-judge-business-school.

European Commission. 2019. "Horizon 2020." Text. Horizon 2020 - European Commission. 2019. https://ec.europa.eu/programmes/horizon2020/en.

European Economic and Social Committee. 2017. "The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context." European Economic and Social Committee. February 22, 2017. https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data.

European Parliament. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)*. *OJ L*. Vol. 119. http://data.europa.eu/eli/reg/2016/679/oj/eng.

Everts, Sarah. 2016. "Information Overload." Science History Institute. July 18, 2016. https://www.sciencehistory.org/distillations/magazine/information-overload.

Floridi, Luciano. 2014. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. OUP Oxford.

Graham, Mark. 2012. "Big Data and the End of Theory?" *The Guardian*, March 9, 2012, sec. News. https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLOS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

Joas, Hans. 1993. *Pragmatism and Social Theory*. University of Chicago Press.

Keulartz, Jozef, Maartje Schermer, Michiel Korthals, and Tsjalling Swierstra. 2004. "Ethics in Technological Culture: A Programmatic Proposal for a Pragmatist Approach." *Science, Technology, & Human Values* 29 (1): 3–29. https://doi.org/10.1177/0162243903259188.

Kostkova, Patty, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, et al. 2016. "Who Owns the Data? Open Data for Healthcare." *Frontiers in Public Health* 4. https://doi.org/10.3389/fpubh.2016.00007.

Kvochko, Elena. 2012. "Four Ways to Talk About Big Data." Text. Information and Communications for Development. December 4, 2012. http://blogs.worldbank.org/ic4d/four-ways-to-talk-about-big-data.

Manyika, James, Michael Chui, Jaques Bughin, and Brad Brown. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity." 2011. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Reprint edition. Boston: Eamon Dolan/Mariner Books.

Mearian, Lucas. 2015. "Insurance Company Now Offers Discounts -- If You Let It Track Your Fitbit." Computerworld. April 17, 2015. https://www.computerworld.com/article/2911594/insurance-company-now-offers-discounts-if-you-let-it-track-your-fitbit.html.

MIKE2.0. 2019. "Big Data Solution Offering - MIKE2.0, the Open Source Methodology for Information Development." 2019. http://mike2.openmethodology.org/wiki/Big_Data_Solution_Offering.

MORO, Veronica. 2016. "Meeting the Challenges of Big Data." Text. European Data Protection Supervisor - European Data Protection Supervisor. November 16, 2016. https://edps.europa.eu/data-protection/our-work/publications/opinions/meeting-challenges-big-data_en.
Nature. 2008. "Community Cleverness Required." *Nature* 455 (7209): 1. https://doi.org/10.1038/455001a.
Noorman, Merel. 2012. "Computing and Moral Responsibility." *Stanford Encyclopedia of Philosophy*.
Norwegian Data Protection Authority. 2013. "Big Data – Privacy Principles under Pressure." https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf.
NSTC (National Science and Technology Council). 2016. "National Privacy Research Strategy." https://obamawhitehouse.archives.gov/sites/default/files/nprs_nstc_review_final.pdf.
OECD. 2013. "New Data for Understanding the Human Condition: International Perspectives." http://www.oecd.org/sti/inno/new-data-for-understanding-the-human-condition.pdf.
———. 2016. "Research Ethics and New Forms of Data for Social and Economic Research," November. https://doi.org/10.1787/5jln7vnpxs32-en.
Ohm, Paul. 2012. "Don't Build a Database of Ruin." *Harvard Business Review*, August 23, 2012. https://hbr.org/2012/08/dont-build-a-database-of-ruin.
Pariser, Eli. 2011. *The Filter Bubble: What The Internet Is Hiding From You*. Penguin Books Limited.
Pels, Dick. 1996. "The Politics of Symmetry." *Social Studies of Science* 26 (2): 277–304. https://doi.org/10.1177/030631296026002004.
Piatetsky, Gregory. 2014. "Interview: Michael Berthold, KNIME Founder, on Research, Creativity, Big Data, and Privacy, Part 2." 2014. https://www.kdnuggets.com/2014/08/interview-michael-berthold-knime-research-big-data-privacy-part2.html, https://www.kdnuggets.com/2014/08/interview-michael-berthold-knime-research-big-data-privacy-part2.html.
Reichman, O. J., Matthew B. Jones, and Mark P. Schildhauer. 2011. "Challenges and Opportunities of Open Data in Ecology." *Science* 331 (February): 703. https://doi.org/10.1126/science.1197962.
Richterich, A. 2018. "The Big Data Agenda: Data Ethics and Critical Data Studies." https://doi.org/10.16997/book14.b.
Rip, Arie. 2013. "Pervasive Normativity and Emerging Technologies." In *Ethics on the Laboratory Floor*, edited by Simone van der Burg and Tsjalling Swierstra, 191–212. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137002938_11.
Snijders, Chris, Uwe Matzat, and Ulf-Dietrich Reips. 2012. "'Big Data': Big Gaps of Knowledge in the Field of Internet Science." http://www.ijis.net/ijis7_1/ijis7_1_editorial.pdf.
Sullivan, Danny. 2015. "Google Still Doing At Least 1 Trillion Searches Per Year." Search Engine Land. January 16, 2015. https://searchengineland.com/google-1-trillion-searches-per-year-212940.

Tanner, Adam. 2014. "Different Customers, Different Prices, Thanks To Big Data." Forbes. 2014. https://www.forbes.com/sites/adamtanner/2014/03/26/different-customers-different-prices-thanks-to-big-data/.

UK Data Service. 2017. "Big Data and Data Sharing: Ethical Issues." https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf.

Zimmer, Michael. 2016. "OkCupid Study Reveals the Perils of Big-Data Science." *Wired*, May 14, 2016. https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/.

Zook, Matthew, Solon Barocas, Danah Boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, et al. 2017. "Ten Simple Rules for Responsible Big Data Research." *PLOS Computational Biology* 13 (3): e1005399. https://doi.org/10.1371/journal.pcbi.1005399.